



Computational Intelligence in Electrical Engineering
Vol. 15, No. 4, 2025
pp. 1-10
Research Paper

Design of A Low-Energy SRAM Cell Capable of Performing In-memory Computations with Application in Neural Networks

Seyed Hassan Hadi Nemati¹, Nima Eslami², Mohammad Hossein Moaiyeri^{*1}

¹ Faculty of Electrical Engineering, Shahid Beheshti University, Tehran, Iran

² Faculty of Electrical Engineering, Shahid Beheshti University, Tehran, Iran

³ Faculty of Electrical Engineering, Shahid Beheshti University, Tehran, Iran

Abstract:

This paper proposed a novel 9T SRAM cell. The key idea is to separate the write and read operations, resolving the inherent conflict in conventional cells like 6T SRAM cells. Moreover, this cell is capable of performing in-memory computation as well. In this design, the data node is floated during the write operation to improve the write margin, ensuring a non-conflicting write operation within the cell. Moreover, to overcome the half-select issue, the design employs both row-based and column-based signaling. The performance of the proposed cell in an actual application is evaluated by simulating the LeNet-5 neural network structure. Based on the computational capabilities of the proposed cell, based on the in-memory AND/NAND and OR/NOR Boolean logic functions and the combination of them to perform XNOR logic, the layers of the binarized LeNet-5 network are implemented after circuit-level simulation. The results demonstrate that the energy consumption of the proposed cell is significantly reduced by up to 46% compared to other existing structures.

Keywords: SRAM cell, Neural Network, In-memory Computing, Energy Consumption.



This is an open access article under the CC BY-NC-ND/4.0/ License (<https://creativecommons.org/licenses/by-nc-nd/4.0/>).



<https://doi.org/10.22108/ISEE.2024.138212.1636>

طراحی سلول حافظه SRAM با انرژی مصرفی کم و قابلیت محاسبه در حافظه برای اجرا

شبکه‌های عصبی

سید حسن هادی نعمتی^۱، نیما اسلامی^۲، محمد حسین معیری^{۳*}

۱- کارشناسی ارشد، دانشکده مهندسی برق، دانشگاه شهید بهشتی، تهران، ایران

s.hadinemati@mail.sbu.ac.ir

۲- دانشجوی دکتری، دانشکده مهندسی برق، دانشگاه شهید بهشتی، تهران، ایران

n_eslami@sbu.ac.ir

۳- دانشیار، گروه الکترونیک، دانشکده مهندسی برق، دانشگاه شهید بهشتی، تهران، ایران

h_moayeri@sbu.ac.ir

چکیده: در این مقاله یک سلول حافظه دسترسی تصادفی ایستا (SRAM) جدید ۹ ترانزیستوری ارائه شده است. در این سلول حافظه، با جداسازی فرآیندهای خواندن از نوشتن، تقابل ذاتی بین این دو فرآیند که در ساختارهای مرسوم وجود دارد، حل شده است. سلول پیشنهادی توانایی اجرای محاسبات در حافظه را دارا است. برای افزایش حاشیه نویز نوشتن در ساختار پیشنهادی، از شیوهٔ شناورسازی گره داده استفاده شده است. این امر سبب جلوگیری از تقابل ترانزیستورها هنگام نوشتن داده در سلول می‌شود. همچنین، به منظور حل مشکل نوشتن ناخواسته در سلول‌های نیمه‌انتخابی، سیگنال‌بندی سطری و ستونی در طرح پیشنهادی در نظر گرفته شده است. برای ارزیابی کاربرد عملکرد سلول پیشنهادی، ساختار شبکهٔ عصبی LeNet-5 دودویی شده با بهره‌گیری از قابلیت محاسباتی سلول SRAM پیشنهادی شبیه‌سازی شده است. در این راستا، با بهره‌گیری از امکان محاسبهٔ منطق‌های AND/NAND و OR/NOR و ترکیب آن‌ها برای اجرای گیت XNOR در ساختار پیشنهادی، لایه‌های متفاوت شبکهٔ LeNet-5 پس از شبیه‌سازی در سطح مداری، به طور کامل به یکدیگر متصل می‌شوند و ساختار اجرا می‌شود. نتایج نشان می‌دهد میزان انرژی مصرفی ساختار پیشنهادی در مقایسه با ساختارهای مقایسه‌شده تا ۶۶ درصد کمتر است.

واژه‌های کلیدی: سلول حافظه SRAM، شبکهٔ عصبی، محاسبه در حافظه، انرژی مصرفی.

۱- مقدمه

کاربردهای نیازمند به حافظه‌های سریع SRAM هستند که در استفادهٔ طولانی‌مدت به مصرف کم‌توان نیاز دارند [۲]. از آنجا که بخشی جالب توجه از مساحت تراشه‌های کامپیوتری توسط این حافظه‌ها اشغال شده است [۳]، کاهش توان مصرفی حافظهٔ SRAM از جمله رویکردهایی است که بسیار مورد توجه پژوهشگران قرار گرفته است. از سوی دیگر، کاهش ابعاد ترانزیستورها سبب بروز چالش‌هایی فراوان همچون اثرات کانال کوتاه، حامل‌های داغ و اشباع سرعت شده است. از جمله رویکردهایی نوین که امروزه مورد استقبال صنعت قرار گرفته‌اند، فناوری جایگزین FinFET است [۴]. در این فناوری، گیت به واسطهٔ ساختار سه‌بعدی خود، از سه وجه بر کانال سیلیکونی احاطه و تا

حافظه‌های SRAM از جمله حافظه‌های پرسرعت هستند که در کاربردهایی وسیع از سیستم‌های کامپیوتری استفاده می‌شوند [۱]. اینترنت اشیا (IoT)، شبکه‌های حسگر بی‌سیم (WSN)، ابزارهای زیست‌پزشکی قابل کاشت و دستگاه‌های الکترونیکی قابل حمل از جمله

^۱ تاریخ ارسال مقاله: ۱۴۰۲/۰۴/۰۶

تاریخ پذیرش مقاله: ۱۴۰۳/۰۷/۰۸

نام نویسندهٔ مسئول: محمد حسین معیری

نشانی نویسندهٔ مسئول: ایران، تهران، دانشگاه شهید بهشتی،

دانشکده مهندسی برق، گروه الکترونیک

در مقاله [۷]، یک ساختار محاسبه در حافظه مبتنی بر سلول حافظه ۸ ترانزیستوری رایج طراحی شده است که توانایی محاسبه عملیات منطقی را دارا است. در مقاله [۸]، یک ساختار ۸ ترانزیستوری ارائه شده است که دارای هسته ۶ ترانزیستوری است و با استفاده از دو ترانزیستور اضافه شده، فرآیند محاسبه در حافظه را روی دو خط داده انجام می‌دهد. در مقاله [۹]، یک ساختار ۹ ترانزیستوری طراحی شده است که با استفاده از هسته ۶ ترانزیستوری و اضافه کردن سه ترانزیستور، فرآیند محاسبه در حافظه را به کمک دو خط داده اجرا می‌کند.

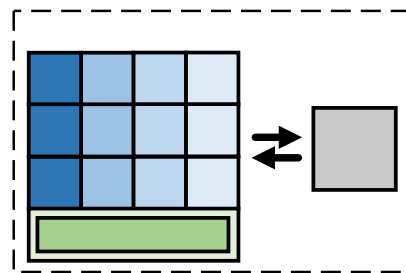
در تمامی ساختارهای محاسبه در حافظه، هسته سلول ساختار ۶ ترانزیستوری رایج است که توانایی نوشتن آن کم است و پایداری مناسبی در این فرآیند نخواهد داشت. همچنین، ساختار ۶ ترانزیستوری رایج در معرض نوشتن ناخواسته در سلول‌های نیمه‌انتخابی است که در ساختارهای محاسبه در حافظه که به محاسبات بیتی در تک‌سلول‌ها نیاز است، این امر میزان مصرف توان سلول‌ها را به دلیل نیاز به بازنویسی داده‌های سلول‌های تخریب‌شده به شدت افزایش خواهد داد [۱۲].

در این مقاله، یک ساختار نوین SRAM مبتنی بر ترانزیستورهای FinFET ارائه شده است که فرآیندهای خواندن و نوشتن را برای اجرای کارآمد محاسبه در حافظه از هم جدا کرده است. همچنین، به منظور بهبود توانایی نوشتن در سلول، ساختار با استفاده از ساختاری نوین و شناورسازی گره ذخیره داده، فرآیند نوشتن را تسهیل کرده است. به منظور حل مشکل نوشتن ناخواسته در سلول‌های نیمه‌انتخابی به منظور افزایش کارایی محاسبه در حافظه و کاهش توان مصرفی، سلول ارائه شده دارای سیگنال‌های فرمان سطری و ستونی است.

در ادامه، در بخش دوم، حافظه پیشنهادی و نحوه عملکرد آن بررسی و به دنبال آن، روش اجرای محاسبات در حافظه بیان می‌شود. در بخش سوم، نتایج شبیه‌سازی‌های انجام‌شده روی سلول بررسی و ارزیابی می‌شود و در بخش پایانی، نتیجه‌گیری مقاله ارائه می‌شود.

حدی بسیار خوب بر جریان عبوری از کانال نظارت دارد [۵]. زدودن ناحیه بالاسر گیت سه‌بعدی ترانزیستور FinFET سبب ایجاد ساختاری با دو گیت مستقل (IG-FinFET) می‌شود [۶]. در این ساختار هر یک از گیت‌ها به طور مستقل قادر به کنترل جریان عبوری کانال است و در صورت فعال شدن یکی از گیت‌ها، ترانزیستور روشن می‌شود.

گفتنی است، با توجه به نیاز روزافزون به کاربردهایی مانند موتورهای جست‌وجو و شبکه‌های عصبی، حجم داده‌های تحت پردازش در حال افزایش است [۷]؛ این در حالی است که با توجه به اینکه معماری بیشتر سیستم‌های کامپیوتری امروزی بر اساس معماری وان‌نیومن است که در آن پردازنده و حافظه مجزا از هم واقع شده‌اند و فقط از طریق گذرگاهی با پهنای باند محدود با یکدیگر ارتباط برقرار می‌کنند، با افزایش حجم اطلاعات تبادل بین پردازنده و حافظه، تأخیر و توان مصرفی سیستم به شدت افزایش خواهد یافت [۸]. به منظور حل این چالش، راهکار محاسبات در حافظه بسیار مورد توجه قرار گرفته است. در این روش بسیاری از محاسبات ساده با کمترین سربار در داخل حافظه اجرا می‌شوند و به تبادل داده بین پردازنده و حافظه نیاز نیست (شکل ۱) [۹].



شکل (۱): ساختار محاسبه در حافظه

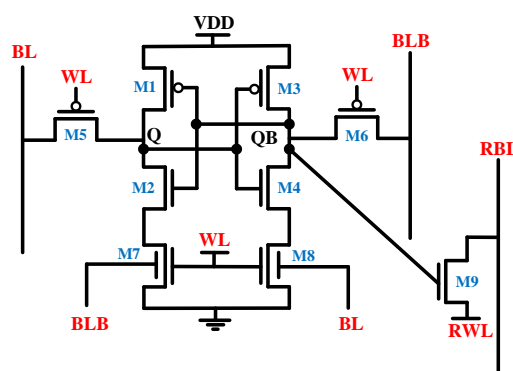
در سال‌های گذشته، ساختارهای گوناگون حافظه SRAM طراحی شده‌اند تا در کنار توانایی ذخیره داده با قابلیت اطمینان زیاد و توان مصرفی کم، توانایی اجرای محاسبات پایه‌ای در حافظه را نیز ممکن کنند. بیشتر ساختارهای ارائه‌شده مبتنی بر ساختار ۶ و ۸ ترانزیستوری رایج هستند [۷]، [۸]، [۹]، [۱۰]، [۱۱].

۲- طرح پیشنهادی

شکل (۲) ساختار ۹ ترانزیستوری پیشنهادی را به تصویر می‌کشد. این ساختار متقارن دارای دو ترانزیستور دسترسی (M5, M6) از نوع pFET برای نوشتن داده روی سلول طراحی شده است. ترانزیستورهای M1, M2, M3 و M4 و M3 و M4 تشکیل‌دهنده دو وارونگر پشت به پشت متصل هستند که بازخورد مثبت آن، داده را درون سلول حفظ می‌کند. همچنین، در این ساختار دو ترانزیستور پشته‌شده M7 و M8 نیز تعبیه شده‌اند تا در فرآیند نوشتن داده، گره مد نظر شناور شود [۱۱] و داده به راحتی داخل سلول قرار داده شود.

۲-۱- حالت نگهداری داده

در این حالت، خطوط داده BL و BLB و RBL به ولتاژ تغذیه متصل هستند. همچنین، سیگنال‌های کنترلی WL و RWL به ولتاژ تغذیه متصل هستند؛ به این ترتیب، داده داخل سلول از محیط بیرون ایزوله و بازخورد مثبت مورد نیاز برای حفظ داده توسط دو وارونگر پشت به پشت برقرار می‌شود. با توجه به اتصال RWL و RBL به ولتاژ تغذیه در حالت نگهداری، میزان جریان‌های نشتی در سلول به میزانی جالب توجه کاهش می‌یابد.



شکل (۲): ساختار طرح پیشنهادی

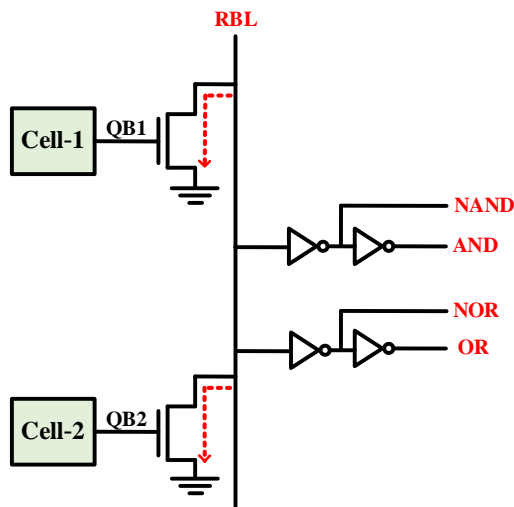
۲-۲- فرآیند خواندن داده

این فرآیند توسط ترانزیستور M9 اجرا می‌شود. در این حالت، به طور پیش‌فرض، خط داده خواندن (RBL) به

ولتاژ تغذیه پیش‌شارژ می‌شود؛ این در حالی است که در حالت اولیه، خط کلمه خواندن (RWL) به ولتاژ تغذیه متصل شده است. با رسیدن فرمان فرآیند خواندن، خط کلمه خواندن از مقدار '1' به مقدار '0' تغییر وضعیت می‌دهد. چنانچه داده موجود در گره QB برابر '1' باشد، خط پیش‌شارژ شده RBL شروع به دشارژ به سمت زمین می‌کند و چنانچه این داده '0' باشد، خط RBL در ولتاژ پیش‌شارژ خود باقی خواهد ماند. این تغییرات به طور هم‌زمان توسط تقویت‌کننده حسی موجود در انتهای آرایه رصد می‌شوند و داده خوانده شده را ثبت می‌کنند. با توجه به بهره‌گیری از یک ترانزیستور در فرآیند خواندن [۸]، سرعت خواندن در این ساختار در مقایسه با سایر طرح‌هایی که مسیر دشارژ با تعداد ترانزیستور بیشتر دارند، زیادتر است.

۲-۳- فرآیند نوشتن داده

در این فرآیند، خط کلمه WL به عنوان سیگنال کنترلی در فرآیند نوشتن ایفای نقش می‌کند. بر این اساس، داده مدنظر ابتدا روی خط داده BL و وارون آن روی خط BLB قرار می‌گیرد و سپس، خط کلمه WL فعال می‌شود ('1' ← '0'). با فعال شدن خط کلمه WL، ترانزیستورهای M5 و M6 روشن می‌شوند و مسیری مستقیم از خطوط داده به گره‌های داده سلول فراهم می‌شود. برای مثال، اگر فرض شود داده '0' درون سلول (گره Q) ذخیره است، چنانچه قرار باشد داده '1' داخل سلول نوشته شود، خط BL به '1' و خط BLB به '0' متصل می‌شود. با فعال شدن خط WL، مقدار '1' از طریق ترانزیستور M5 به گره Q منتقل می‌شود. هم‌زمان با این فرآیند، از آنجا که خط BLB حاوی مقدار '0' است و خط WL نیز به '0' متصل شده است، ترانزیستور M7 خاموش است و عملاً گره Q شناور است و داده '1' بدون هیچ رقابت و تقابلی با سرعت زیاد درون سلول نوشته خواهد شد. در حالت دیگر، چنانچه درون سلول داده '1' ذخیره شده باشد، برای نوشتن داده



'0' درون این سلول، نیاز است تا خطوط BL و BLB به '0' و '1' به ترتیب متصل شوند. سپس، با فعال کردن خط WL، مقادیر موجود روی خطوط داده به گره‌های داده سلول منتقل می‌شوند.

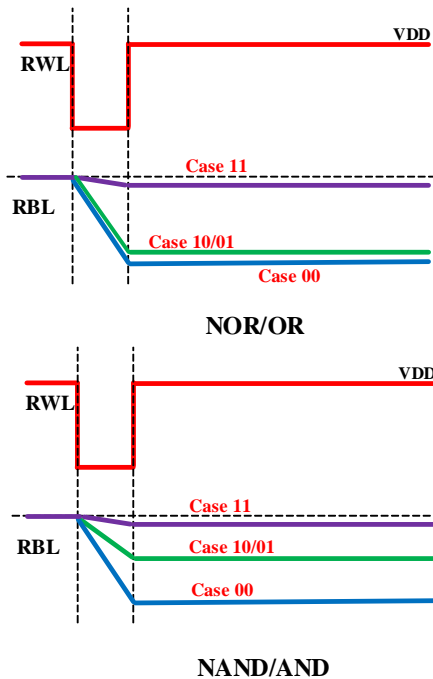
نکته حائز اهمیت در سلول پیشنهادی آن است که در صورتی که فقط یکی از خطوط BL یا BLB یا WL تغییر وضعیت دهد، هیچ کدام از گره‌های داده به حالت شناوری نمی‌رود و بازخورد درون سلول برقرار خواهد ماند تا از نوشتن ناخواسته در سلول‌های نیمه‌انتخابی جلوگیری شود.

۴-۲- ساختار محاسبه در حافظه

در ساختار پیشنهادی، با جداسازی فرآیند خواندن از نوشتن، توانایی اجرای محاسبه در حافظه ایجاد شده است. در این راستا، ساختار پیشنهادی توانایی اجرای محاسبات بولین را در ستون دارا است. به منظور اجرای محاسبه OR/NOR طبق شکل (۳)، ابتدا خط RBL به میزان ولتاژ تغذیه شارژ اولیه می‌شود.

سپس، در گام دوم، بر اساس داده ذخیره‌شده در سلول حافظه اول و دوم، خط RBL تخلیه می‌شود یا بدون تغییر باقی خواهد ماند. برای مثال، اگر داده ذخیره‌شده در گره QB یکی از سلول‌ها یا هر دو '1' باشد، خط RBL تخلیه می‌شود؛ این در حالی است که اگر این داده در هر دو سلول '0' باشد، خط RBL دست‌نخورده باقی خواهد ماند. به منظور اجرای AND/NAND، نیز مانند OR/NOR همین فرآیند اجرا می‌شود؛ با این تفاوت که میزان زمان فعال‌سازی RWL محدود می‌شود.

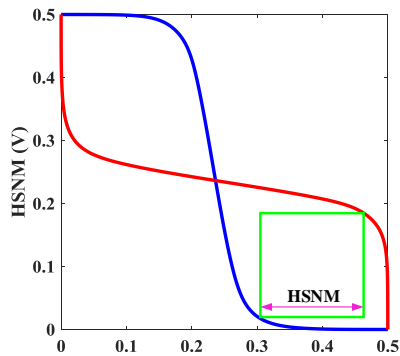
در این شرایط، خط RBL زمانی به صورت کامل تخلیه می‌شود که داده گره QB در هر دو سلول '1' باشد؛ در غیر این صورت، خط RBL به میزان کافی تخلیه نخواهد شد.



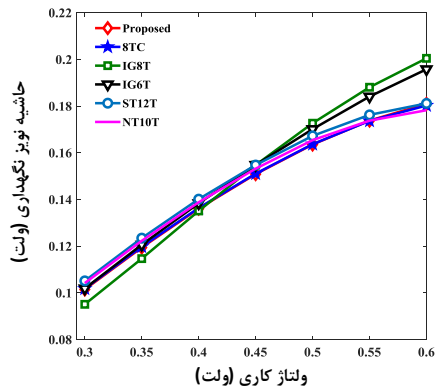
شکل (۳): محاسبه AND/NAND و OR/NOR

۳- نتایج شبیه‌سازی و ارزیابی سلول پیشنهادی

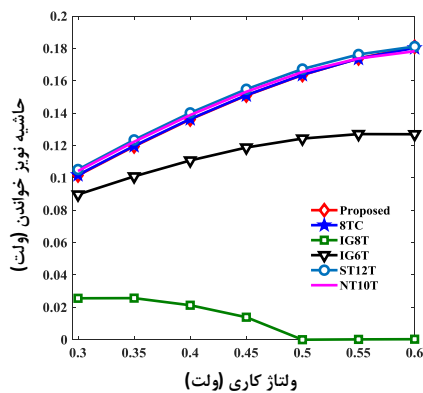
به منظور ارزیابی و شبیه‌سازی سلول پیشنهادی، از فناوری فین فت گیت مستقل (IG-FinFET) استفاده شده است [۱۳]، [۱۴]. برخی از پارامترهای مهم این فناوری در جدول (۱) مشاهده می‌شوند. نکته حائز اهمیت در این شبیه‌سازی‌ها آن است که به منظور ارزیابی منصفانه در تمام سلول‌های مقایسه‌شده از فین‌فت‌های تک‌باله استفاده شده است. برای ارزیابی ساختار پیشنهادی، مقایسه‌ها با طرح‌های



الف) نمودار پروانه ای



ب) حاشیه نویز نگهداری



پ) حاشیه نویز خواندن

شکل (۴): الف) نمودار پروانه ای، ب) حاشیه نویز نگهداری،

پ) حاشیه نویز خواندن در ولتاژهای تغذیه گوناگون

۳-۲- حاشیه نویز ایستای خواندن (RSNM)

همانند معیار HSNM، معیار حاشیه نویز ایستای خواندن با اندازه‌گیری بزرگ‌ترین ضلع مربعی محاسبه می‌شود که داخل کوچک‌ترین ناحیه نمودار پروانه‌ای در فرآیند خواندن ایجاد می‌شود [۳]. از آنجا که با بهره‌گیری از مسیر مجزای خواندن، فرآیند خواندن داده از سلول پیشنهادی، هیچ تاثیری روی گره‌های ذخیره داده ندارد،

موجود به‌روز (8TC [۷]، IG8T [۶]، IG6T [۱۵])،
ST12T [۱۶] و NT10T [۲] انجام شده‌اند.

جدول (۱): برخی از پارامترهای مهم فناوری IG-FinFET

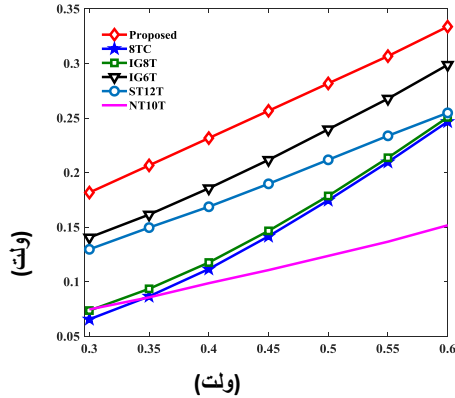
پارامترها	مقادیر (هر دو گیت)
پهنای باله	۸ nm
طول گیت	۳۰ nm
ضخامت اکسیدگیت معادل	۱,۴ nm / ۱,۴ nm
تابع کار گیت	۴,۶۱ eV / ۴,۶۱ eV
دوپینگ کانال	$۱۰^{۲۲} \text{ cm}^{-3}$
غلظت دوپینگ سورس/درین	$۲^{۲۶} \text{ cm}^{-3}$
دوپینگ بدنه	$۵^{۲۳} \text{ cm}^{-3}$

۱-۳- حاشیه نویز ایستای نگهداری (HSNM)

حاشیه نویز ایستای نگهداری معیاری معتبر و شناخته‌شده برای سنجش میزان مقاومت سلول در برابر نویزی است که هدف تخریب داده را در حالت نگهداری دارد. این معیار با اندازه‌گیری بزرگ‌ترین ضلع مربع محاسبه می‌شود که داخل کوچک‌ترین ناحیه نمودار پروانه قابل محاط است [۱۶]. شکل (۴-الف) نمایشی از نمودار پروانه-ای ساختار پیشنهادی در ولتاژ کاری ۰/۵ ولت را به تصویر می‌کشد. به منظور ارزیابی بهتر سلول، مقایسه‌ای میان HSNM طرح پیشنهادی با پنج ساختار ارائه‌شده در سال‌های گذشته انجام شده است (شکل ۴-ب). همان‌طور که در شکل (۴-ب) مشاهده می‌شود، HSNM سلول پیشنهادی بسیار نزدیک به سایر طرح‌های ارائه‌شده است و در حفظ داده مانند سایر طرح‌ها عمل می‌کند.

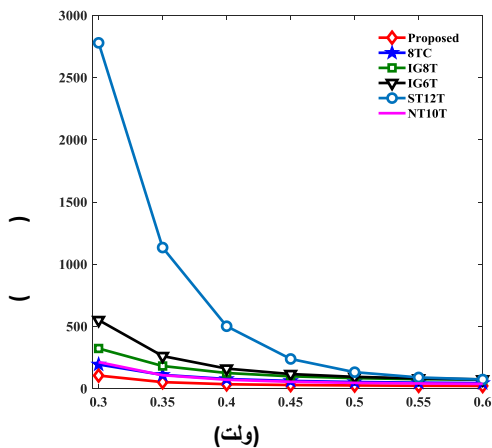
طراحی سلول حافظه SRAM با انرژی مصرفی کم و قابلیت محاسبه در حافظه برای اجرا شبکه‌های عصبی

دلیل استفاده از سه ترانزیستور در مسیر خواندن، سرعتی بسیار کمتر نسبت به سایر طرح‌ها دارد.



شکل (۵): نوشت پذیرگی سلول پیشنهادی در ولتاژهای گوناگون

یکی دیگر از پارامترهای مهم در ارزیابی عملکرد سلول‌های SRAM میزان انرژی مصرفی در فرآیند خواندن است. نتایج شبیه‌سازی محاسبه انرژی در شکل (۷) نشان می‌دهد طرح پیشنهادی کارایی مشابهی ساختارهای 8TC، NT10T و IG6T را از خود نشان می‌دهد و ساختار ST12T به واسطه تأخیر زیاد خواندن، دارای انرژی مصرفی خواندن زیادی نسبت به سایر طرح‌های مقایسه‌شده است.



شکل (۶): تأخیر خواندن سلول پیشنهادی در ولتاژهای گوناگون

۳-۵- تأخیر و انرژی مصرفی در فرآیند نوشتن

همان‌طور که شکل (۸) نشان می‌دهد، تأخیر نوشتن

مقادیر RSNM و HSNM برابر خواهند بود. شکل (۴-پ) مقایسه RSNM طرح پیشنهادی را با سایر طرح‌ها نشان می‌دهد.

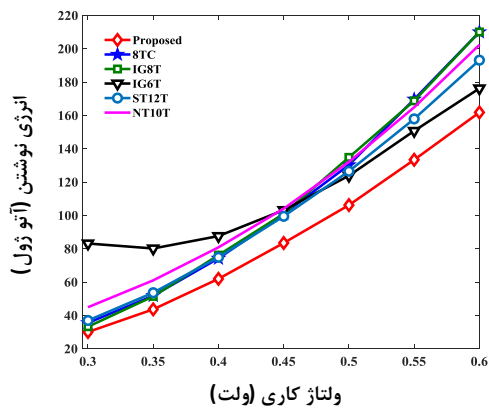
همان‌طور که مشاهده می‌شود، طرح پیشنهادی حاشیه نویز خواندن بسیار زیادی نسبت به دو طرح IG8T و IG6T دارد. نکته حائز اهمیت در این مقایسه آن است که طرح IG8T در فرآیند خواندن از سلول با مشکل نیمه-انتخابی مواجه است و داده سلول در این فرآیند تغییر خواهد کرد. این امر سبب شده است تا حاشیه نویز خواندن از سلول در ولتاژهای زیاد نزدیک به صفر شود.

۳-۳- قابلیت نوشتن داده

ارزیابی فرآیند نوشتن در پژوهش‌های مختلف بر مبنای معیارهایی مختلف انجام می‌شود. از جمله شناخته‌شده‌ترین این معیارها میزان قابلیت نوشت پذیرگی داده در سلول (CWLM) است [۱۷]. این معیار با اجرای شبیه‌سازی حالت ایستا و تغییر WL از '1' به سمت '0' محاسبه می‌شود تا فاصله ولتاژ WL نسبت به زمین، حاشیه نویز نوشتن را به تصویر کشد (صرفاً برای ترانزیستور نوع pFET) [۱۸]. شکل (۵) نوشت پذیرگی سلول پیشنهادی را در قیاس با سایر طرح‌ها نشان می‌دهد. شناورسازی گره ذخیره داده همراه با نوشتن به صورت دیفرانسیلی سبب افزایش جالب توجه نوشت پذیرگی سلول پیشنهادی می‌شود. همان‌طور که مشاهده می‌شود، ساختار دیفرانسیلی IG8T و ساختار 8TC به دلیل وجود رقابت در مدار بالابر با ترانزیستورهای دسترسی، نوشت پذیرگی کمی دارند.

۳-۴- تأخیر و انرژی مصرفی در فرآیند خواندن

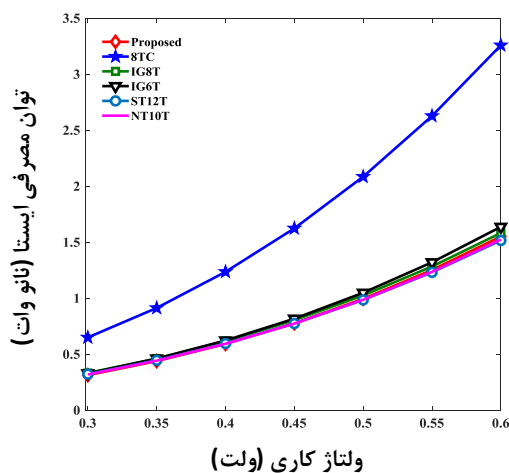
شکل (۶) مقایسه تأخیر خواندن طرح پیشنهادی را با سایر طرح‌های مقایسه‌شده نشان می‌دهد. همان‌طور که در این شکل نشان داده شده است، ساختار پیشنهادی کمترین زمان خواندن را دارا است. علت سرعت زیاد ساختار پیشنهادی، تک‌ترانزیستور بودن مدار خواندن سلول است که مقاومتی بسیار کم را برای تخلیه RBL از خود نشان می‌دهد. در میان طرح‌های مقایسه‌شده، ساختار ST12T به



شکل (۹): انرژی مصرفی نوشتن سلول پیشنهادی در ولتاژهای گوناگون

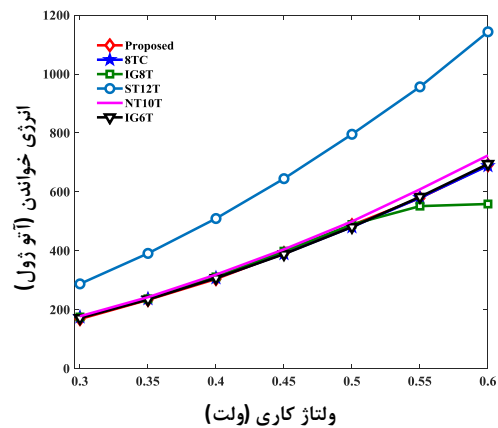
۶-۳- توان مصرفی ایستا

شکل (۱۰) توان مصرفی ایستای سلول‌ها را در ولتاژهای کاری مختلف نشان می‌دهد. طرح پیشنهادی در ولتاژهای تغذیه پایین، توان مصرفی کمتری در قیاس با سایر طرح‌های مقایسه شده دارد و در ولتاژهای تغذیه بالا جای خود را به ساختار ST12T و NT10T خواهد داد. این امر بدان علت است ساختارهای بیان شده ترانزیستورهای پشته خاموش بیشتری در حالت نگهداری دارند. گفتنی است، سلول 8TC بیشترین توان مصرفی ایستا را نسبت به سایر طرح‌ها دارا است که دلیل اصلی آن قرارگیری خطوط BL و BLB سلول در حالت پیش‌شارژ است که به افزایش نشتی از ولتاژ تغذیه به مسیرهای متاهی به زمین منجر می‌شود.

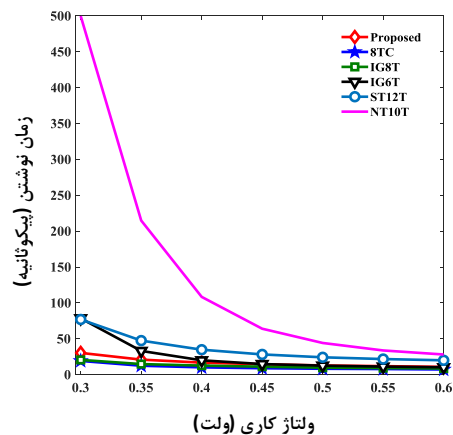


شکل (۱۰): توان مصرفی ایستای سلول پیشنهادی در ولتاژهای گوناگون

سلول پیشنهادی در مقایسه با طرح‌هایی همچون ST12T که فقط از روش شناورسازی برای نوشتن داده بهره برده‌اند، عملکردی بهتر را به ارمغان آورده است. همچنین، ساختار IG6T که ترانزیستورهای دسترسی آن مقاومتی زیاد دارند و به دنبال آن، طرح NT10T، عملکردی ضعیف‌تر نسبت به طرح پیشنهادی دارند. از سوی دیگر، شکل (۹) به خوبی نشان می‌دهد طرح‌های 8TC و IG8T که از زمان نوشتن خوبی برخوردار هستند، به دلیل توان مصرفی زیاد، انرژی مصرفی بیشتری را نسبت به طرح پیشنهادی دارا هستند. با توجه به شکل (۹)، مشاهده می‌شود طرح پیشنهادی دارای کمترین انرژی مصرفی در فرآیند نوشتن است.



شکل (۷): انرژی مصرفی خواندن سلول پیشنهادی در ولتاژهای گوناگون



شکل (۸): تأخیر نوشتن سلول پیشنهادی در ولتاژهای گوناگون

۷-۳- ارزیابی کاربردی ساختار پیشنهادی

برای ارزیابی ساختار ارائه‌شده به صورت کاربردی، ساختار بالا در اجرای شبکه عصبی دودویی بررسی و ارزیابی شده است. شبکه عصبی دودویی جایگزینی برای شبکه‌های عصبی با دقت کامل است تا میزان توان مصرفی و حجم محاسبات را با کاهش اندک در دقت حاصل کند [۱۹] به طور کلی، در شبکه‌های عصبی بیشترین سربار توان مربوط به لایه کانولوشن است [۲۰]؛ از این رو، به منظور اجرای این لایه در شبکه عصبی دودویی، نیاز است تا عملیات XNOR به جای ضرب پنجره داده در وزن مربوط در هر لایه انجام شود و به دنبال آن، پس از محاسبات بالا، جمع یک‌های حاصل در مرحله قبل به کمک جمع‌کننده محاسبه شود [۲۱]. برای مثال، در اجرای لایه کانولوشن به کمک پنجره داده با اندازه ۳×۳، نیاز است تا این پنجره ورودی روی تمام لایه کانولوشن حرکت کند و در هر سیکل، عدد باینری وزن لایه در داده داخل پنجره، XNOR شده و به دنبال آن، مجموع تعداد نتایج یک‌شده در محاسبات XNOR با هم جمع شود.

از این رو، در گام نخست، به منظور اجرای XNOR از چهارگیت NAND استفاده شده است؛ از این رو، محاسبه XNOR داده در وزن برای هر داده نیازمند چهار سیکل محاسبات است. شکل (۱۱) ساختار کلی شبکه عصبی استفاده‌شده و محاسبات مربوط به XNOR را در داخل حافظه به تصویر می‌کشد.

در این مدل، در ابتدا، محاسبات مربوط به XNOR در حافظه برای داده ورودی و وزن دودویی در نرم‌افزار HSPICE شبیه‌سازی و پارامترهای تأخیر و انرژی مصرفی ساختار پیشنهادی استخراج می‌شوند. سپس، با استفاده از نرم‌افزار MATLAB، ساختار شبکه عصبی اجرا می‌شود و مقادیر به‌دست‌آمده از مرحله اجرای مدار بر روی شبکه طراحی شده مدل می‌شوند [۲۲]. شبکه استفاده‌شده در این بخش شبکه LeNet-5 و مجموعه داده استفاده‌شده برای ارزیابی سیستم اجراشده، مجموعه داده Fashion-MNIST

است [۲۳]، [۲۴]. نکته جالب توجه آن است که شبکه LeNet-5 استفاده‌شده به صورت از پیش آموزش‌دیده برای کار با داده‌های ورودی و وزن دودویی انتخاب شده است. ساختار پیشنهادی با سه ساختار نوین محاسبه در حافظه مقایسه شده است. این سه ساختار جزء ساختارهایی هستند که توانایی اجرای محاسبه در حافظه را دارا هستند.

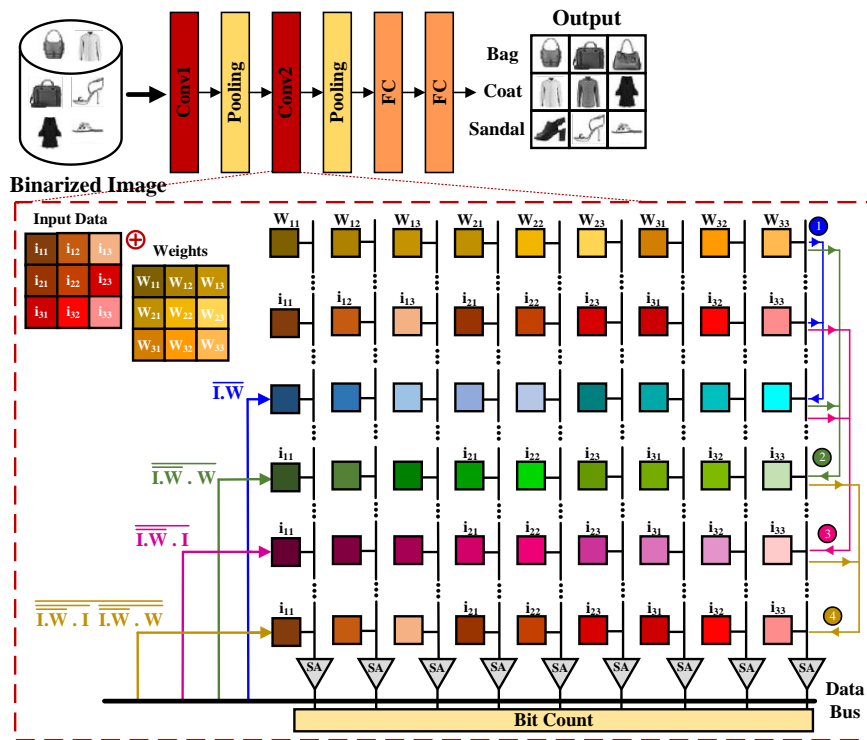
جدول (۲) میزان تأخیر و انرژی مصرفی شبکه مدنظر را نشان می‌دهد. با توجه به جدول (۲)، مشاهده می‌شود انرژی مصرفی و تأخیر ساختار پیشنهادی از سایر ساختارهای مقایسه‌شده کمتر هستند. این امر به آن علت است که در ساختارهای مقایسه‌شده در محاسبات مربوط به XNOR، تأخیر و انرژی مصرفی در فرآیند نوشتن داده‌های میانی و خواندن آن‌ها برای محاسبات پیش‌رو از ساختار پیشنهادی بیشتر هستند. همچنین، تمامی ساختارهای مقایسه‌شده از مشکل نوشتن ناخواسته در سلول‌های نیمه‌انتخابی رنج می‌برند که سبب می‌شود تا مصرف توان در این ساختارها به دلیل نیاز به نوشتن مجدد داده در سلول‌های تخریب‌شده به شدت افزایش یابد.

میزان انرژی مصرفی در اجرای ساختار بالا بر اساس سلول پیشنهادی در مقایسه با ساختارهای ST12T، 8TC و NT10T به ترتیب به میزان ۱۸ درصد، ۴۶ درصد و ۲۲ درصد کمتر است. همچنین، گفتنی است، میزان تأخیر اجرای شبکه تحت بررسی نسبت به ساختارهای بیان‌شده به میزان ۳۷ درصد، ۷۵ درصد و ۵۷ درصد به ترتیب کمتر می‌باشد.

جدول (۲): تأخیر و انرژی مصرفی شبکه عصبی LeNet-5

دودویی شده

ساختار حافظه	تأخیر (نانوثانیه)	انرژی مصرفی (میکرو ژول)
8TC	۱۲۳٫۴	۲٫۹۲
ST12T	۳۱۱٫۵۶	۴٫۴۱
NT10T	۱۸۱	۳٫۰۲
پیشنهادی	۷۷٫۶	۲٫۳۸



شکل (۱۱): ساختار شبکه عصبی اجرا شده و مراحل محاسبات XNOR

۴- نتیجه گیری

8TC، ST12T و NT10T به ترتیب به میزان ۳۷ درصد، ۷۵ درصد و ۵۷ درصد کمتر بوده است. همچنین، میزان انرژی مصرفی در اجرای شبکه عصبی تحت بررسی در مقایسه با ساختارهای بیان شده به ترتیب ۱۸ درصد، ۴۶ درصد و ۲۲ درصد کمتر است.

مراجع

- [1] E. Abbasian, "A Highly Stable Low-Energy 10T SRAM for Near-Threshold Operation", IEEE Transactions on Circuits and Systems I: Regular Papers, Vol. 69, No. 12, 2022. doi: 10.1109/TCSI.2022.3207992.
- [2] E. Abbasian, S. Sofimowloodi, "Energy-Efficient Single-Ended Read/Write 10T Near-Threshold SRAM", IEEE Transactions on Circuits and Systems I: Regular Papers, Vol. 70, No. 5, 2023. doi: 10.1109/TCSI.2023.3247807.
- [3] S. Ahmad, N. Alam, M. Hasan, "Pseudo differential multi-cell upset immune robust SRAM cell for ultra-low power applications", AEU - International Journal of Electronics and Communications, Vol. 83, pp. 366-375, 2018. doi: 10.1016/j.aeue.2017.09.022.
- [4] S. Sayyah Ensan, M. H. Moaiyeri, M. Moghaddam, S. Hessabi, "A Low-Power Single-Ended SRAM in FinFET Technology", International Journal of Electronics and

در این مقاله، یک سلول SRAM ۹ ترانزیستوری با مصرف توان کم طراحی شده است. در این ساختار، برای حل مشکل تقابل بین فرآیندهای خواندن و نوشتن، مسیرهای این دو فرآیند از هم جدا شده‌اند. همچنین، ساختار پیشنهادی قابلیت اجرای محاسبات بیتی OR/NOR و AND/NAND در حافظه را فراهم می‌کند. در این ساختار، به منظور افزایش حاشیه نویز نوشتن، گره ذخیره داده شناور می‌شود تا داده به راحتی داخل سلول قرار داده شود. همچنین، به منظور حل مشکل نوشتن ناخواسته در سلول‌های نیمه‌انتخابی، فرمان‌های سطری و ستونی اعمال می‌شوند تا گره ذخیره داده در سلول‌های نیمه‌انتخابی شناور نشود. در ساختار ارائه شده، میزان توان مصرفی نوشتن ۱۹ درصد کمتر از ساختار ۸ ترانزیستوری رایج در اجرای محاسبات در حافظه است. به منظور ارزیابی کاربردی ساختار پیشنهادی، این ساختار در اجرای شبکه عصبی LeNet-5 دودویی شده ارزیابی شده است. نتایج شبیه‌سازی نشان می‌دهد میزان تأخیر سیستم نسبت به ساختارهای

- Nanoscale Integration and Modeling (NIMO) Group. [Online]. Available: <http://ptm.asu.edu/>
- [15] R. Niaraki Asli, S. Taghipour, "Reliable and high performance asymmetric FinFET SRAM cell using back-gate control", *Microelectronics Reliability*, Vol. 104, p. 113545, Jan. 2020. doi: 10.1016/J.MICROREL.2019.113545.
- [16] M. Karamimanesh, E. Abiri, K. Hassanli, M. R. Salehi, A. Darabi, "A robust and write bit-line free sub-threshold 12T-SRAM for ultra-low power applications in 14 nm FinFET technology", *Microelectronics J*, Vol. 118, 2021. doi: 10.1016/j.mejo.2021.105185.
- [17] N. Eslami, B. Ebrahimi, E. Shakouri, D. Najafi, "A single-ended low leakage and low voltage 10T SRAM cell with high yield", *Analog Integr Circuits Signal Process*, Vol. 105, No. 2, 2020. doi: 10.1007/s10470-020-01669-y.
- [18] H. Makino, Sh. Nakata, H. Suzuki, Sh. Mutoh, M. Miyama, T. Yoshimura, "Reexamination of SRAM cell write margin definitions in view of predicting the distribution", *IEEE Transactions on Circuits and Systems II: Express Briefs*, Vol. 58, No. 4, pp. 230–234, 2011. doi: 10.1109/TCSII.2011.2124531.
- [19] C. H. Wang, K. Y. Huang, Y. Yao, J. C. Chen, H. H. Shuai, W. H. Cheng, "Lightweight Deep Learning: An Overview", *IEEE Consumer Electronics Magazine*, Vol. 13, No. 4, 2024. doi: 10.1109/MCE.2022.3181759.
- [20] S. M. Rizvi, A. A. H. A. Rahman, U. U. Sheikh, K. A. A. Fuad, H. M. F. Shehzad, "Computation and memory optimized spectral domain convolutional neural network for throughput and energy-efficient inference", *Applied Intelligence*, Vol. 53, No. 4, 2023. doi: 10.1007/s10489-022-03756-1.
- [21] S. Angizi, M. Morsali, S. Tabrizchi, A. Roohi, "A Near-Sensor Processing Accelerator for Approximate Local Binary Pattern Networks", *IEEE Trans Emerg Top Comput*, Vol. 12, No. 1, 2024. doi: 10.1109/TETC.2023.3285493.
- [22] N. Eslami, M. H. Moaiyeri, "A Flexible and Reliable RRAM-Based In-Memory Computing Architecture for Data-Intensive Applications", *IEEE Trans Emerg Top Comput*, 2023. doi: 10.1109/TETC.2023.3268079.
- [23] G. Xu, M. Liu, Z. Jiang, W. Shen, C. Huang, "Online Fault Diagnosis Method Based on Transfer Convolutional Neural Networks", *IEEE Trans Instrum Meas*, Vol. 69, No. 2, 2020. doi: 10.1109/TIM.2019.2902003.
- [24] Y. Ying, J. Su, P. Shan, L. Miao, X. Wang, S. Peng, "Rectified Exponential Units for Convolutional Neural Networks", *IEEE Access*, Vol. 7, 2019. doi: 10.1109/ACCESS.2019.2928442.
- Communications*, Vol. 99, pp. 361–368, 2018, [Online]. Available: <https://www.sciencedirect.com/science/article/abs/pii/S1434841118312615>
- [5] F. Salmanpour, M. H. Moaiyeri, F. Sabetzadeh, "Ultra-Compact Imprecise 4:2 Compressor and Multiplier Circuits for Approximate Computing in Deep Nanoscale", *Circuits Syst Signal Process*, Vol. 40, No. 9, 2021. doi: 10.1007/s00034-021-01688-8.
- [6] S. Sayyah Ensan, M. H. Moaiyeri, B. Ebrahimi, S. Hessabi, A. Afzali-Kusha, "A low-leakage and high-writable SRAM cell with back-gate biasing in FinFET technology", *J Comput Electron*, Vol. 18, pp. 519–526, 2019. doi: 10.1007/s10825-019-01327-1.
- [7] A. Agrawal, A. Jaiswal, C. Lee, K. Roy, "X-SRAM: Enabling in-memory Boolean computations in CMOS static random-access memories", *IEEE Transactions on Circuits and Systems I: Regular Papers*, Vol. 65, No. 12, pp. 4219–4232, 2018. doi: 10.1109/TCSI.2018.2848999.
- [8] Y. Chen, J. Mu, H. Kim, L. Lu, T. T. H. Kim, "BP-SCIM: A Reconfigurable 8T SRAM Macro for Bit-Parallel Searching and Computing In-Memory", *IEEE Transactions on Circuits and Systems I: Regular Papers*, Vol. 70, No. 5, 2023. doi: 10.1109/TCSI.2023.3240303.
- [9] K. Soundrapandiyan, S. K. Vishvakarma, B. S. Reniwal, "Enabling Energy-Efficient In-Memory Computing with Robust Assist-Based Reconfigurable Sense Amplifier in SRAM Array", *IEEE J Emerg Sel Top Circuits Syst*, Vol. 13, No. 1, 2023. doi: 10.1109/JETCAS.2023.3243192.
- [10] Z. Lin, H. Zhan, X. Li, Ch. Peng, W. Lu X. Wu, "In-Memory Computing with Double Word Lines and Three Read Ports for Four Operands", *IEEE Trans Very Large Scale Integr VLSI Syst*, Vol. 28, No. 5, pp. 1316–1320, 2020. doi: 10.1109/TVLSI.2020.2976099.
- [11] S. Mittal, G. Verma, B. Kaushik, F. A. Khanday, "A survey of SRAM-based in-memory computing techniques and applications", 2021. doi: 10.1016/j.sysarc.2021.102276.
- [12] S. H. H. Nemati, N. Eslami, M. H. Moaiyeri, "A Hybrid SRAM/RRAM In-Memory Computing Architecture Based on a Reconfigurable SRAM Sense Amplifier", *IEEE Access*, Vol. 11, 2023. doi: 10.1109/ACCESS.2023.3294675.
- [13] F. Kenarangi I. Partin-Vaisband, "Leveraging Independent Double-Gate FinFET Devices for Machine Learning Classification", *IEEE Transactions on Circuits and Systems I: Regular Papers*, Vol. 66, No. 11, 2019. doi: 10.1109/TCSI.2019.2927441.
- [14] "Predictive Technology Model (PTM)",

¹ Internet of Things

² Wireless Sensor Networks